

Exam

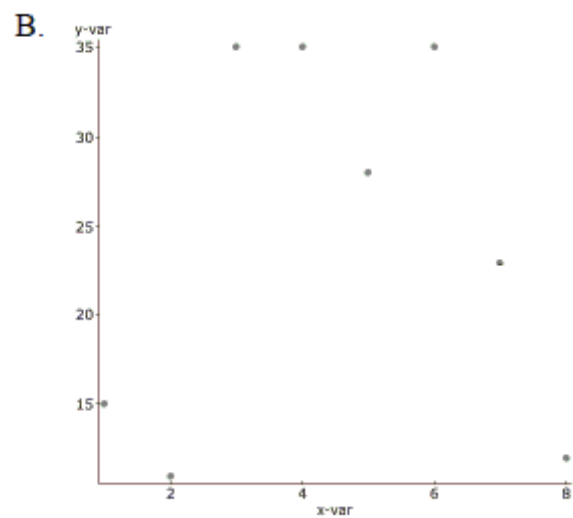
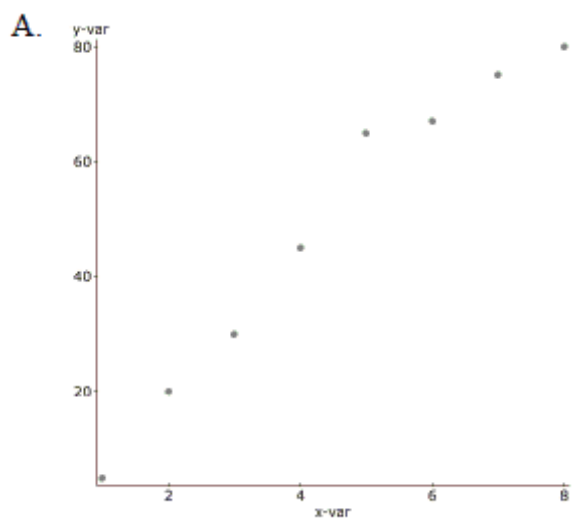
Name _____

SHORT ANSWER. Write the word or phrase that best completes each st

1) Match each graph with the appropriate correlation coefficient.

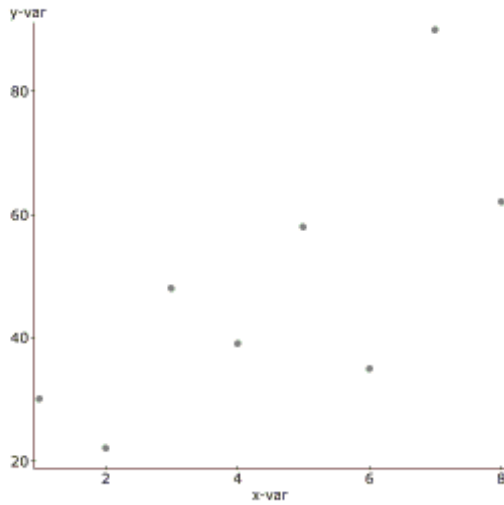
_____ 0.98 _____ 0.73 _____ 0.09 _____ -0.99

1) _____

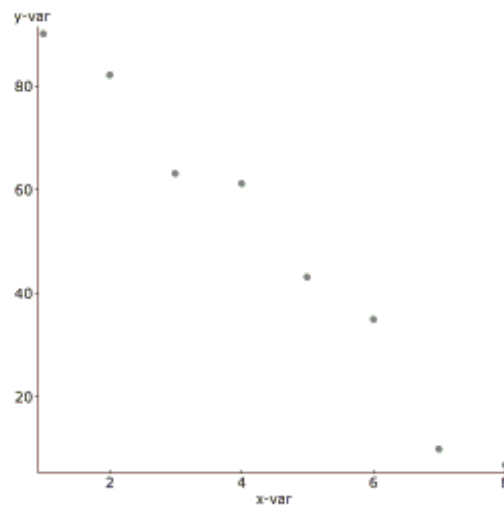


C.

1)



D.



2) After conducting a survey of his students, a professor reported that "There appears to be a strong correlation between grade point average and whether or not a student works." Comment on this observation.

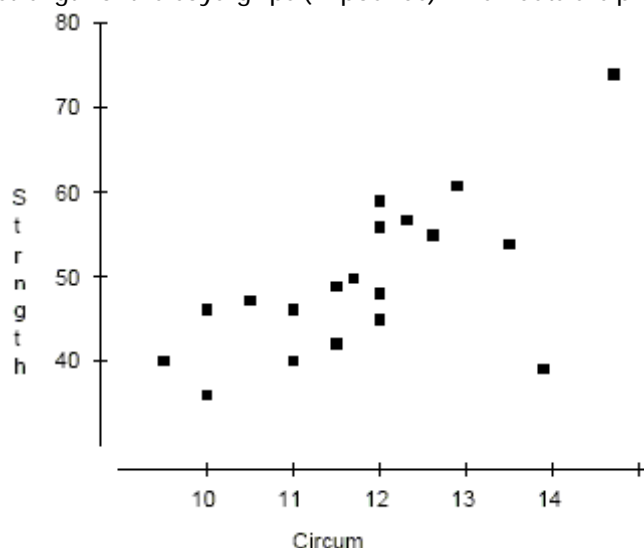
2) _____

3) After conducting a marketing study to see what consumers thought about a new tinted contact lens they were developing, an eyewear company reported, "Consumer satisfaction is strongly correlated with eye color." Comment on this observation.

3) _____

- 4) Researchers investigating the association between the size and strength of muscles measured the forearm circumference (in inches) of 20 teenage boys. Then they measured the strength of the boys' grips (in pounds). Their data are plotted below.

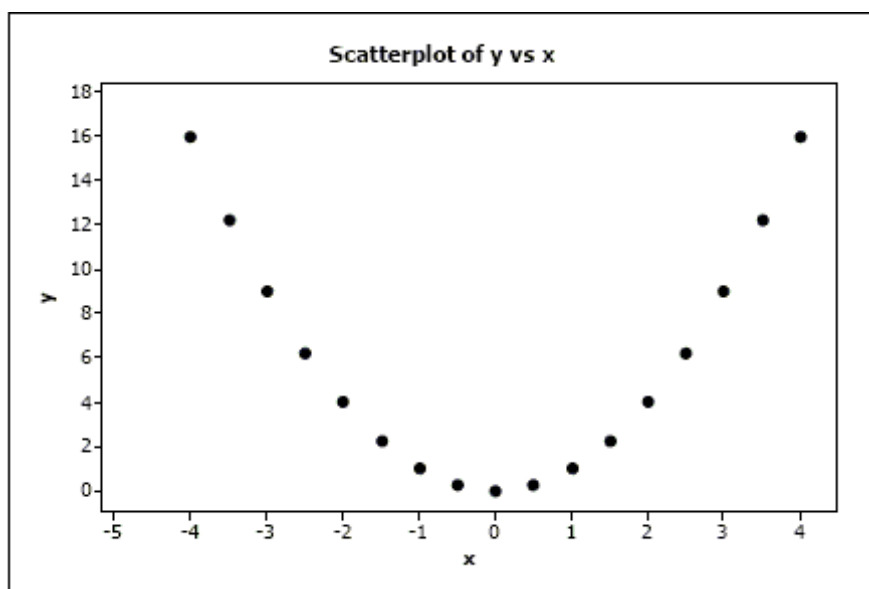
4) _____



- Write a few sentences describing the association.
- Estimate the correlation. $r =$ _____
- If the point in the lower right corner (at about 14" and 38 lbs.) were removed, would the correlation become stronger, weaker, or remain about the same?
- If the point in the upper right corner (at about 15" and 75 lbs.) were removed, would the correlation become stronger, weaker, or remain about the same?

- 5) The following scatterplot shows a relationship between x and y that results in a correlation coefficient of $r = 0$. Explain why $r = 0$ in this situation even though there appears to be a strong relationship between the x and y variables.

5) _____



In an effort to decide if there is an association between the year of a postal increase and the new postal rate for first class mail, the data were gathered from the United States Postal Service. In 1981, the United States Postal Service changed their rates on March 22 and November 1. This information is shown in the table below.

Year	Rate
1971	0.08
1974	0.10
1975	0.13
1978	0.15
1981	0.18
1981	0.20
1985	0.22
1988	0.25
1991	0.29
1995	0.32

- 6) Would it be better for customers for a year to have a negative residual or a positive residual from this model? Explain.

6) _____

An article in the *Journal of Statistics Education* reported the price of diamonds of different sizes in Singapore dollars (SGD). The following table contains a data set that is consistent with this data, adjusted to US dollars in 2004:

2004 US \$	Carat	2004 US \$	Carat	2004 US \$	Carat
494.82	0.12	688.24	0.15	748.10	0.16
768.03	0.17	944.90	0.18	1076.18	0.19
1105.03	0.20	1071.75	0.21	1289.20	0.23
1508.88	0.25	1504.44	0.26	1597.63	0.27
1826.18	0.28	1908.28	0.29	2038.09	0.32
2096.89	0.33	2409.76	0.35		

- 7) Make a scatterplot and describe the association between the size of the diamond (carat) and the cost (in US dollars).

7) _____

- 8) Gas mileage An important factor in the amount of gasoline a car uses is the size of the engine. Called "displacement", engine size measures the volume of the cylinders in cubic inches. The regression analysis is shown.

8) _____

Dependent variable is: MPG
 89 total cases of which 0 are missing
 R squared = 60.9% R squared (adjusted) = 60.0%
 s = 3.056 with 89 - 2 = 87 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	696.744	1	696.744	74.6
Residual	448.236	48	9.33826	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	34.9799	1.231	28.4	< 0.0001
Eng. Displcmt	-0.066196	0.0077	-8.64	< 0.0001

- How many cars were included in this analysis?
- What is the correlation between engine size and fuel economy?
- A car you are thinking of buying is available with two different size engines, 190 cubic inches or 240 cubic inches. How much difference might this make in your gas mileage? (Show your work.)

A study examined the number of trees in a variety of orange groves and the corresponding number of oranges that each grove produces in a given harvest year. Linear regression was calculated and the results are below.

linear regression results:

Dependent Variable: oranges

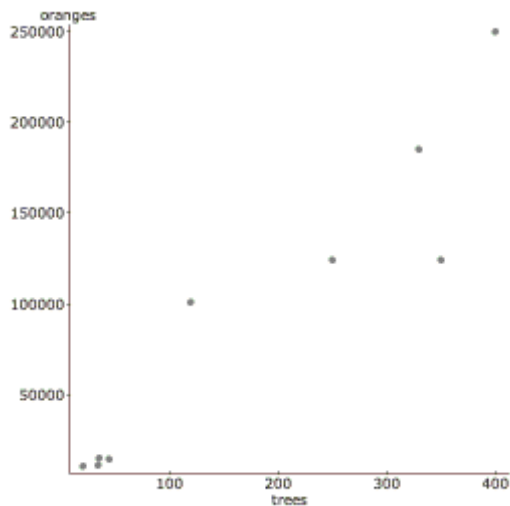
Independent Variable: trees

Sample size: 9

R-sq = 0.886

s = 31394.7

Parameter	Estimate	Std. Err.
Constant	390.59	16328.8
Trees	525.84	71.22



9) Interpret the slope in context.

9) _____

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

10) A silly psychology student gathers data on the shoe size of 30 of his classmates and their GPA's.

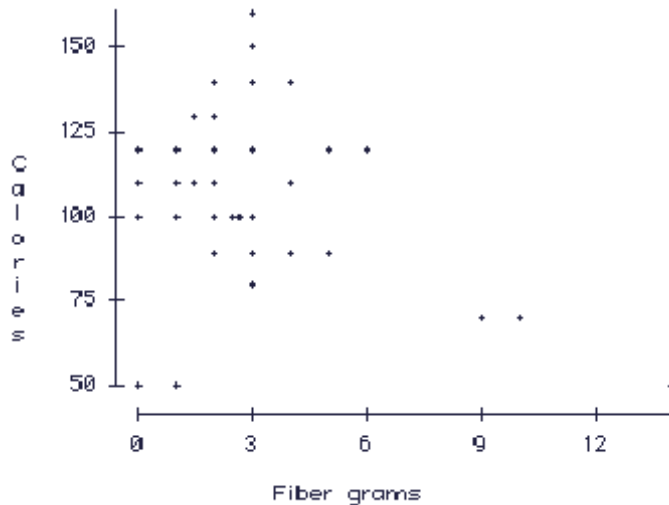
10) _____

The correlation coefficient between these two variables is most likely to be

- A) near 0
- B) exactly +1.0
- C) near +0.6
- D) near -0.6
- E) exactly -1.0

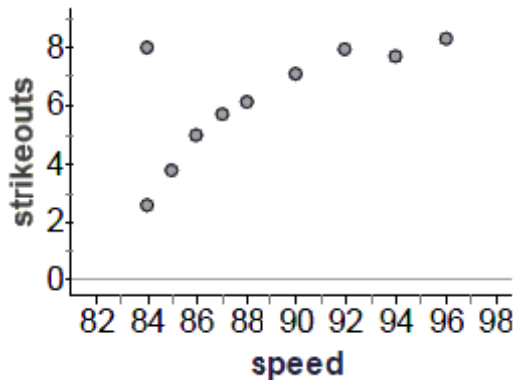
SHORT ANSWER. Write the word or phrase that best completes each st

Current research states that a good diet should contain 20-35 grams of dietary fiber. Research also states that each day should start with a healthy breakfast. The nutritional information for 77 breakfast cereals was reviewed to find the grams of fiber and the number of calories per serving. The scatterplot below shows the relationship between fiber and calories for the cereals.



11) Do you think there is a clear pattern? Describe the association between fiber and calories. 11) _____

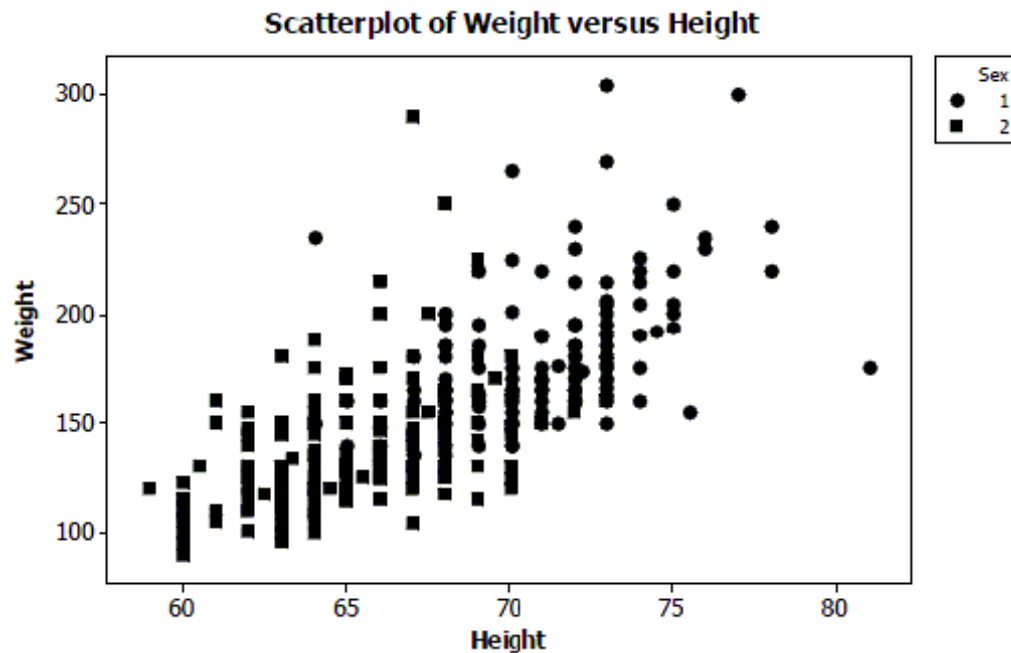
Baseball coaches use a radar gun to measure the speed of pitcher's fastball. They also record outcomes such as hits and strikeouts. The scatterplot below shows the relationship between the average speed of a fastball and the average number of strikeouts per nine innings for each pitcher on the Bulldogs, based on the past season.



12) Do you think the association would be stronger or weaker if we used data from one month of the season? 12) _____

- 13) Here is a scatterplot of weight versus height for students in an introductory statistics class. The men are coded as '1' and appear as circles in the scatterplot; the women are coded as '2' and appear as squares in the scatterplot.

13) _____



- Do you think there is a clear pattern? Describe the association between weight and height.
- Comment on any differences you see between men and women in the plot.
- Do you think a linear model from the set of all data could accurately predict the weight of a student with height 70 inches? Explain.

Halloween is a fun night. It seems that older children might get more candy because they can travel further while trick-or-treating. But perhaps the youngest kids get extra candy because they are so cute. Here are some data that examine this question, along with the regression output.

Dependent Variable: candy

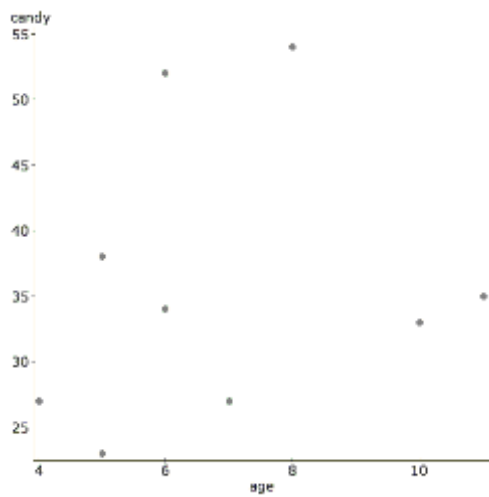
Sample size: 9

R (correlation coefficient) = 0.19534425

R-sq = 0.038159375

s = 11.297554

Parameter	Estimate	Std. Err.
Intercept	29.767327	12.21138
Age	0.88861386	1.6862229



- 14) The next day, a young girl reveals that her older brother also went trick-or-treating, but didn't want to admit that he participated. He was added to the data set and these are the results.

Dependent Variable: candy

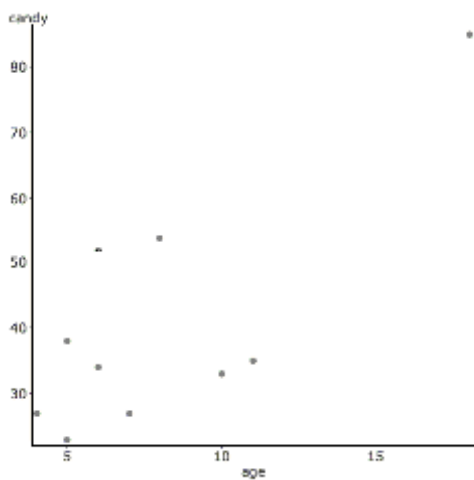
Sample size: 10

R (correlation coefficient) = 0.76362369

R-sq = 0.58312115

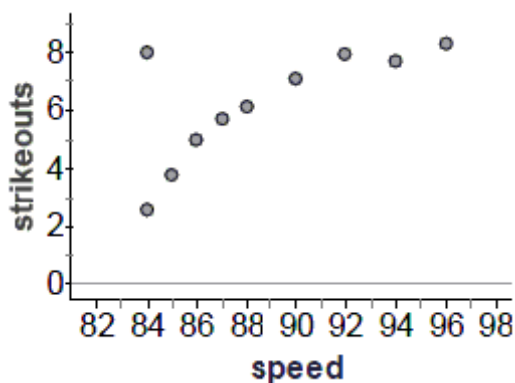
s = 12.709041

Parameter	Estimate	Std. Err.
Intercept	13.569231	9.0783516
Age	3.4038462	1.0175376



Describe the effect of this new candy collector on the regression model.

Baseball coaches use a radar gun to measure the speed of pitcher's fastball. They also record outcomes such as hits and strikeouts. The scatterplot below shows the relationship between the average speed of a fastball and the average number of strikeouts per nine innings for each pitcher on the Bulldogs, based on the past season.



- 15) Do you think a model based on these data could accurately predict the average number of strikeouts for a pitcher with an average fastball speed of 70 mph.? Explain.

15) _____

During a science lab, students heated water, allowed it to cool, and recorded the temperature over time. They computed the difference between the water temperature and the room temperature. The results are in the table.

Time (in minutes)	10	20	30	40	50	60
Difference in temp. (degrees F)	68	36	20	10	6	4

16) Use the equation to predict the difference in temperature after 45 minutes.

16) _____

17) You are given the following costs to build a square deck for your house:

17) _____

Width (ft)	4	5	6	7	8	9
Cost (\$)	150	255	350	500	650	800

Width (ft)	10	11	12	13	14	15
Cost (\$)	1000	1200	1450	1700	1950	2250

- Use re-expressed data to create a model that predicts the cost of the deck based on the width.
- Why do you think that your model is appropriate?
- Find the predicted cost of a square deck that is 10.5 feet wide.
- Is it reasonable to use this model to predict the cost of a square deck that is 20 feet wide? Explain.

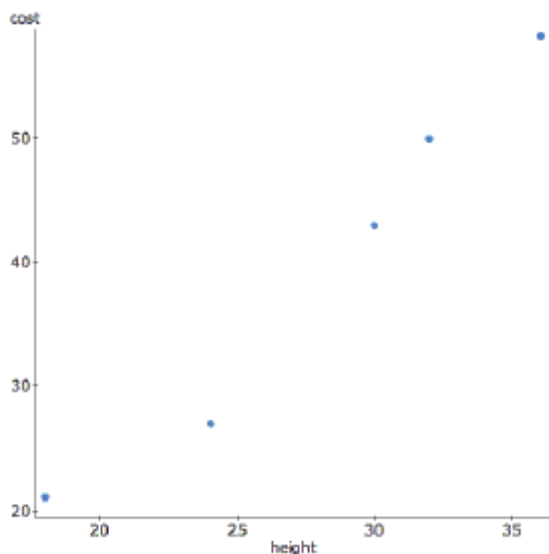
18) During a chemistry lab, students were asked to study a radioactive element which decays over time. The results are in the table.

18) _____

Time (in days)	0	2	4	6	8	10
Element (in grams)	320	226	160	115	80	57

- Model the remaining mass of the element.
- Find the predicted amount of the element remaining after thirty minutes.

The bigger the stop sign, the more expensive it is. Here is a graph of the height of a sign in inches versus its cost in dollars.



To achieve linearity, the data was transformed using a square root function of cost. Here are the results and a residual

plot.

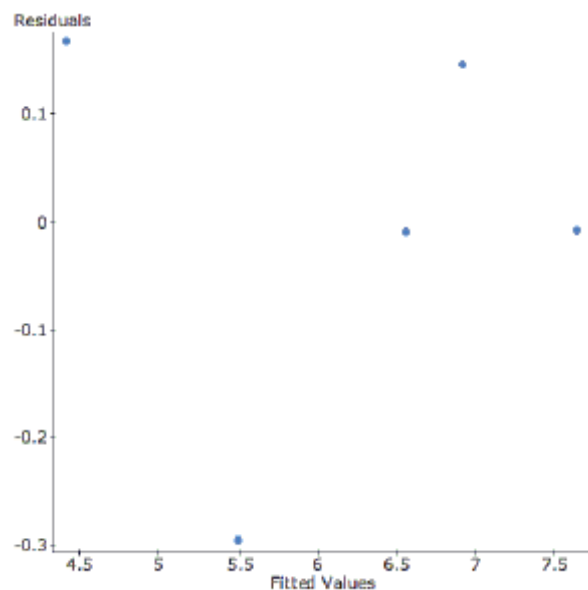
Dependent Variable: $\sqrt{\text{cost}}$

R (correlation coefficient) = 0.98946627

R-sq = 0.97904349

s: 0.2141

Parameter	coeff	se
Intercept	1.1857	0.4346
height	0.1792	0.0151



19) Interpret R-sq in the context of this problem.

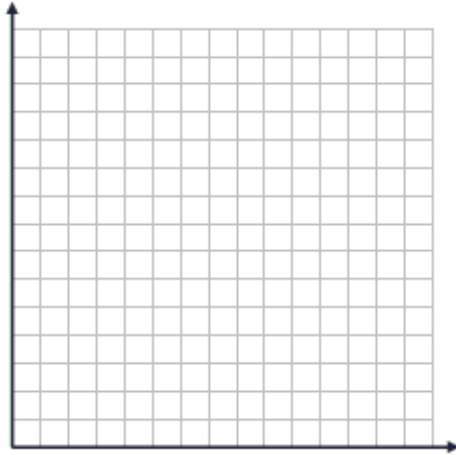
19) _____

During a science lab, students heated water, allowed it to cool, and recorded the temperature over time. They computed the difference between the water temperature and the room temperature. The results are in the table.

Time (in minutes)	10	20	30	40	50	60
Difference in temp. (degrees F)	68	36	20	10	6	4

20) Sketch a scatterplot.

20) _____



MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

21) In predicting the growth of the volume of a small bay by measuring the height of the water at a dock, a researcher is using a model of $\sqrt[3]{\hat{volume}} = 2.34 + 4.56(height)$, where height is measured in m and volume cubic miles. If the height rises to 3.45 m, what is the predicted volume? 21) _____

- A) $7 \times 10^7 \text{ m}^3$
- B) 18.1 m^3
- C) 2.62 m^3
- D) $1.2 \times 10^{18} \text{ m}^3$
- E) 5902 m^3

22) A business owner notes that for every extra hour his store is open, his total sales increase by a fixed amount. His most useful predictive model is probably..... 22) _____

- A) linear
- B) power
- C) logarithmic
- D) exponential
- E) quadratic

23) R-sq is a measure of 23) _____

- A) the probability that the regression line makes a correct prediction.
- B) the change in the y-variable that corresponds with the change in the x-variable.
- C) the proportion of the variability in the response variable that is explained by the explanatory variable.
- D) the initial predicted starting point of the response variable when x is zero.
- E) the percentage of the accuracy of the regression equation.

SHORT ANSWER. Write the word or phrase that best completes each st

24) Mistakes. Describe the mistake made in the following analyses:

24) _____

- a. Ten teachers compute their average test scores for all their students. Then the superintendent collects their data and finds the school average. He repeats this process for eight different schools and finds a positive correlation between the age of the school average age of the teachers at a school and their average score.
- b. The mayor of a city is concerned that the population of the city is growing faster than revenue. He calculates that over the last 5 years, the year and the size of the city have a R-sq of 95.7%. With such a high value, the mayor confidently predicts the population for the next three years of fiscal planning.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

25) Although there are annual ups and downs, over the long run, growth in the stock market averages about 9% per year. A model that best describes the value of a stock portfolio is probably:

25) _____

- A) power
- B) linear
- C) exponential
- D) quadratic
- E) logarithmic